

# Linguistic stability across the lifespan: implications for forensic authorship analysis

Emily Chiang (e.chiang2@aston.ac.uk), Annina Heini (a.heini14@aston.ac.uk), Krzysztof Kredens (k.j.kredens@aston.ac.uk)

## STUDY BACKGROUND

The (in)stability of individuals' linguistic choices across the lifespan becomes an issue for forensic authorship analysis in cases where comparative language samples occur across a large amount of time.

Example scenario: a prolific child sexual offender convicted 20 years ago appears to have resurfaced on the dark web, identity obscured. How do we know whether and to what extent their language choices from 20 years ago remain reliable markers of their linguistic style today?

Improved understanding of individual linguistic stability over the lifespan can help us better evaluate the reliability of forensic authorship analysis methods in cases involving texts written decades apart.

## DATA AND METHODS

**Data:** a 75,054-word corpus of transcripts from the *Up* television series documenting the lives of 14 individuals every seven years between the ages of 7 and 63.

**Methods:** average zeta scores determined the 10 most significant unigrams for the 10 highest-contributing speakers. From this word list we extracted up to three unigrams per speaker judged to be genuinely idiolectal (removing heavily context-dependent words such as names of individuals significant to participants). We noted frequencies of these unigrams at each age point for each participant and conducted a concordance analysis to examine contextual meanings, functions, and usage patterns.

## CONCLUSIONS

- Most individuals show some consistent selection of particular forms from a wide pool of alternatives, demonstrating some level of stability across the lifespan
- No particular feature types appear more stable than others (although stance adverbials make up a large proportion of the idiolectal unigrams)
- In general, usage of idiolectal terms increased with age, suggesting these habitual choices become more entrenched over time (Figure 1)

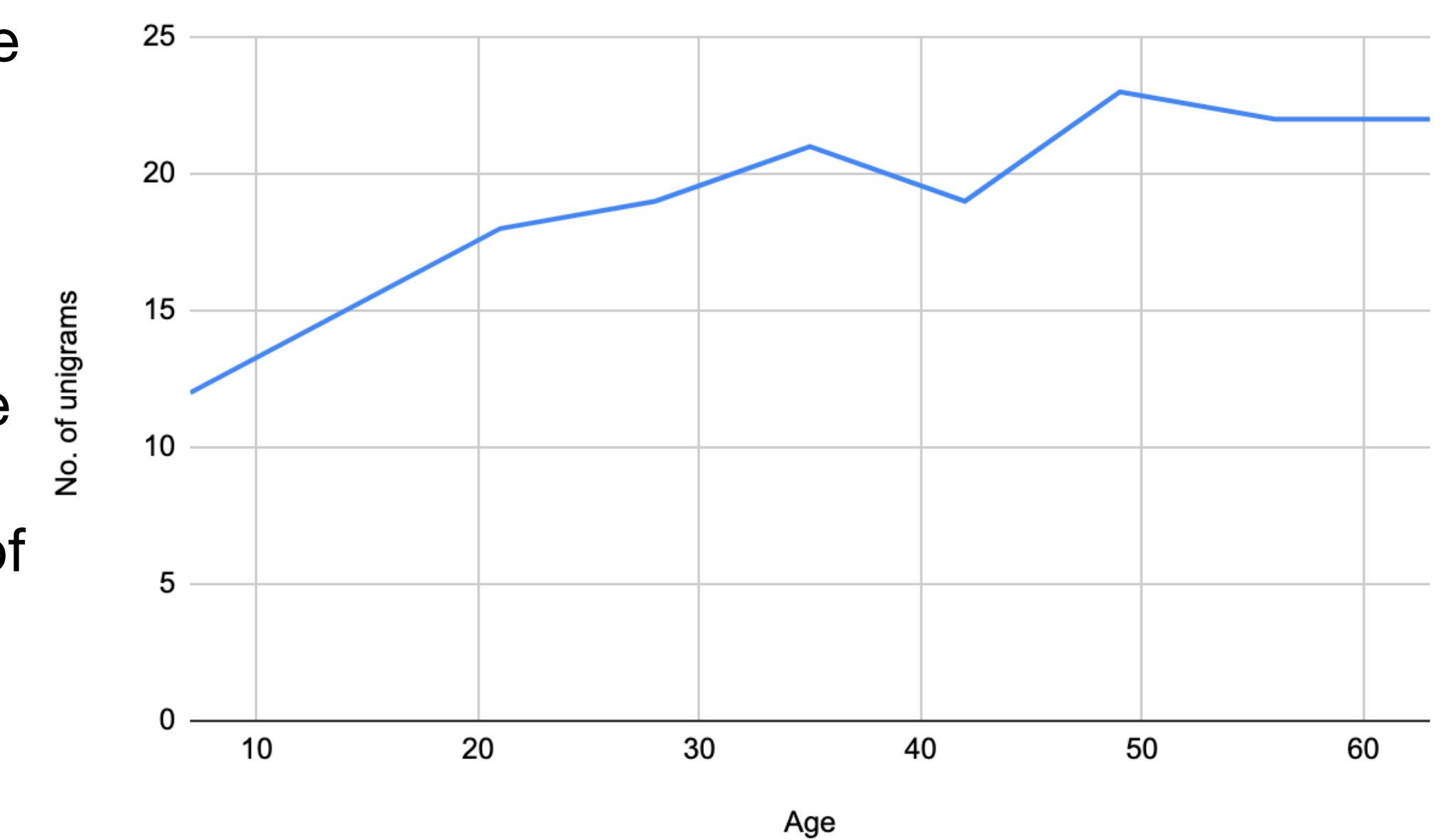












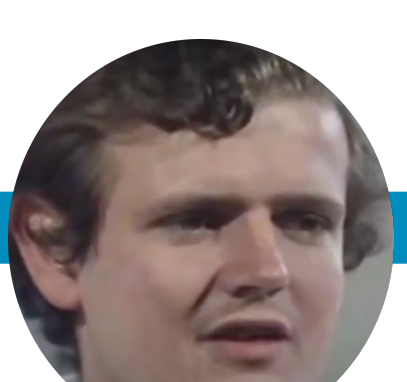
















Figure 1. Number of unigrams observed at each age point for the ten speakers

## ANALYSIS (THREE EXAMPLES)

	7	14	21	28	35	42	49	56	63
<b>Neil:</b> <i>found, perhaps, possibly</i>	 found	 found <b>perhaps</b>	 found possibly	 found <b>perhaps</b> possibly	 found <b>perhaps</b> possibly	 found <b>perhaps</b>	 found <b>perhaps</b> possibly	 found <b>perhaps</b> possibly	 found <b>perhaps</b> possibly
<b>Bruce:</b> <i>quite, or, suppose</i>	 quite <b>or</b>	 <b>or</b> suppose	 quite <b>or</b> suppose	 quite <b>or</b>	 quite <b>or</b> suppose	 quite <b>or</b> suppose	 quite <b>or</b>	 quite <b>or</b> suppose	 quite <b>or</b> suppose
<b>Jackie:</b> <i>thing, probably, best</i>	 thing best	 thing <b>probably</b>	 thing <b>probably</b>	 thing <b>probably</b>	 thing <b>probably</b> best	 thing <b>probably</b> best	 thing <b>probably</b> best	 thing <b>probably</b> best	 thing <b>probably</b> best

**Neil's 'perhaps'**

- predominantly used to explore aspects of his own identity
- reflects a general tendency for introspection
- refers to Neil himself in earlier years, shifting towards others later in life

... *Perhaps I'm not mature enough* (age 7)  
*Perhaps they see the impact on their lives* (age 49)

**Bruce's 'or'**

- predominantly used to coordinate the same/similar ideas, or a specific idea with something indeterminate
- creates the impression Bruce likes to 'cover all bases'
- contributes to a non-committal, hesitant speaking style

... *I'm a little innocent at time **or** naïve* (age 28)  
*I haven't got married **or** whatever* (age 35)

**Jackie's 'probably'**

- predominantly used to reflect on Jackie's own past and character
- consistently selected over lower-certainty adverbs (perhaps, maybe, etc.)
- contributes to a more confident speaking style compared to Neil and Bruce

... *He **probably** knew me better than I did* (age 28)  
*I **probably** should have done further education* (age 63)